EDITORA IBERO-AMERICANA DE EDUCAÇÃO

Revista Hipótese

MUSIC PRODUCTION BY VOICE COMMANDS IN A SOUND PERCEPTION APPLICATION

PRODUÇÃO MUSICAL POR COMANDOS DE VOZ EM UM APLICATIVO DE PERCEPÇÃO SONORA

PRODUCCIÓN MUSICAL POR COMANDOS DE VOZ EN UNA APLICACIÓN DE PERCEPCIÓN DE SONIDO

Leonardo Porto PASSOS
State University of Campinas (UNICAMP)
e-mail: leoportopassos@gmail.com

José FORNARI
State University of Campinas (UNICAMP)
e-mail: fornari@unicamp.br

How to refer to this article

PASSOS, L. P. FORNARI, J. Music production by voice commands in a sound perception application. **Revista Hipótese**, Bauru, v. 8, esp. 1, e022020, 2022. e-ISSN: 2446-7154. DOI: https://doi.org/10.47519/eiaerh.v8.2022.ID414

Submitted: 10/03/2022
Revisions required: 05/05/2022
Approved: 01/07/2022
Published: 01/12/2022

***ABSTRACT**:* In this article, we present the prototype of a web application accessible to the visually impaired for music production education and training in sound perception. Through voice commands, the user can mix the music being played by adding and removing musical instruments and audio effects and changing panning. Mixing is of fundamental importance in the musical recording process, but many listeners and musicians are unaware of the tools and techniques used in this stage, not being able to identify such procedures, which give uniqueness and special aesthetic characteristics to a recording, which motivated the development of the proposed app.

KEYWORDS: Music production. Sound perception. Application development.

***RESUMO**: Neste artigo, apresentamos o protótipo de um aplicativo web acessível a deficientes visuais para educação em produção musical e treinamento em percepção sonora, com o qual o usuário pode, por meio de entradas por comandos de voz, mixar a música em execução, ao adicionar e remover instrumentos musicais e efeitos de áudio e alterar o panning. A mixagem é de fundamental importância no processo de gravação musical, mas muitos ouvintes e músicos desconhecem as ferramentas e técnicas utilizadas nessa etapa, não conseguindo identificar tais procedimentos, que conferem singularidade e características estéticas especiais a uma gravação, o que motivou o desenvolvimento do app proposto.*

***PALAVRAS-CHAVE**: Produção musical. Percepção sonora. Desenvolvimento de aplicativo.*

***RESUMEN**: En este artículo presentamos el prototipo de una aplicación web accesible para personas con discapacidad visual para la educación en producción musical y la formación en percepción del sonido, con la cual el usuario puede, a través de comandos de voz, mezclar la música en ejecución, agregando y quitando instrumentos musicales. y efectos de audio y cambiar la panorámica. La mezcla es de fundamental importancia en el proceso de grabación musical, pero muchos oyentes y músicos desconocen las herramientas y técnicas que se emplean en esta etapa, no pudiendo identificar dichos procedimientos, los cuales le dan singularidad y características estéticas especiales a una grabación, lo que motivó el desarrollo de la aplicación propuesta.*

***PALABRAS CLAVE**: Producción musical. Percepción del sonido. Desarrollo de aplicaciones.*

Introduction

It is very common for a listener, especially a non-musician, to have a special predilection for a certain musical composition without knowing exactly why this affection, considering there is something in that music that evokes his emotions but is beyond his comprehension. And when the same music is appreciated in a live performance, without the use of the same studio resources, that affective relationship is often broken, and the feeling is not the same, even though the performance was very faithful to the musical recording cherished by that person. What may occur in these cases is that the listener has some appreciation for the aesthetics imbued by the music producer to that recording (not counting the participation of the composer and the performer, since our focus here is the musical production, more precisely, the mixing), giving it new contours (beyond those offered by composers and performers), that are often difficult to be reproduced during the performance, for several reasons, ranging from differences in the acoustic treatment of the recording studio environment about the performance venue to the equipment available for sound collection, recording, processing, and generation.

In the words of David Huron (2015, p. 1, our translation), “Music can evoke a wide range of feeling states, from the trivial to the sublime.",[1] this occurs because of four types of emotional generators: 1) association: “[...] certain sounds or sound patterns can be associated with past emotional experiences.”;[2] 2) empathic: "[...] the listener recognizes acoustic features associated with particular emotions.”;[3] 3) cognitive: “Conscious thoughts may lead the listener to a particular experience.”;[4] and 4) signaling: “[...] a signal that changes the observer's behavior.”[5] (HURON, 2012, p. 479, our translation).

In addition to the roles of the composer and performer in evoking feelings in the listener, there is also the music producer's participation in the mixing process, as Richard James Burgess (2013, p. 73) states: “Mixing extends all the musical techniques that precede it, strengthening the perception of the music by reinforcing structure, orchestration, and emotional affect for the intended audience.”.[6] Mixing can intensify the evocation of emotion, or contribute to it, through the proper use of equalizer, filters, distortion, chorus, dynamics, compressor, reverb, echo, pitch shifting, etc. (CASE, 2011), and concerns the following process:

[1] “*Music is capable of evoking a wide range of feeling states from the pedestrian to the sublime*.”
[2] “[…] *certain sounds or sound patterns may become associated with past emotional experiences*.”
[3] “[…] *a listener recognizes acoustic features associated with particular emotions*.”
[4] “*Conscious thoughts can lead a listener to a particular experience*.”
[5] “[…] *a signal is to change the behavior of the observer*.”
[6] “*Mixing extends all the musical techniques that precede it, strengthening the perception of the song by reinforcing the structure, orchestration, and emotional affect for the intended audience*.”

DOI: https://doi.org/10.47519/eiaerh.v8.2022.ID414

> [...] refers to the original mix of a track when the instrumentation and vocals are balanced against each other, and any necessary effects or treatments are added. [...] A mix should optimize the music, vocals, performances, arrangement, and engineering. It should sound good on a wide range of high-end and low-end systems and at any volume (BURGESS, 2013, p. 102, our translation).

Music production - which includes a sound recording, arranging, orchestration, effects, mixing, mastering, etc., as we will see in more detail below - is fundamental for it to be possible to obtain quality in the capture of instruments and voices and for everything to be heard with clarity and definition, according to various aesthetic concepts, which may even favor certain "imperfections" and low-fidelity[7] (*low-fidelity* ou *lo-fi*). (lo-fi) sound. There is great concern with the clarity of musical textures[8] and the separation of the parts in a process that will make the recording procedures imperceptible, or at least reduce the methods of music production (capturing and recording, adding effects, mixing, mastering) so that the recording is perceived as a faithful or "real" representation of the musical performance (TURINO, 2008).

However, as Burgess (2013, p. 2, our translation) states, the processes and techniques of music production, as well as its outcomes, are unknown to many people, even musicians: "[...] I think music production is a poorly understood art, even in the industry."[9] [musical]. And it is to offer the experience of performing a basic musical mix that the prototype of a web[10] application for training in music perception was developed, and thus useful for music production education since it allows the user to understand certain resources and techniques used by music producers in mixing and creative conceptions in aesthetic terms.

To this end, we will explain below what music production is, the definition of music perception, and the reasons that led us to choose the development of an application with inputs by voice commands.

[7] Available at: https://ora.ox.ac.uk/objects/uuid:cc84039c-3d30-484e-84b4-8535ba4a54f8. Access: 10 Jan. 2022.

[8] "The term texture refers to the way melodic, rhythmic, and harmonic materials are woven together in a composition" (*The term texture refers to the way the melodic, rhythmic, and harmonic materials are woven together in a composition*) (BENWARD; SAKER, 2009, p. 145, our translation).

[9] "[…] *I felt that the art of music production was poorly understood, even within the industry*".

[10] Available at: https://edu.gcfglobal.org/pt/informatica-basica/o-que-e-um-aplicativo-web/1/.

DOI: https://doi.org/10.47519/eiaerh.v8.2022.ID414

Music Production

The history of music production began with the emergence of recording, reproduction, and media, attributed to Thomas Alva Edison with the invention of his Phonograph in 1877, a device for recording and reproducing sounds from a cylinder, which was configured as a conceptual and aesthetic milestone of music production by enabling, in fact, the "solidification", so to speak, of the intangibility of the sound material that makes up music, allowing all forms of sound processing and analysis. This new apparatus and the consequent development of new technologies for sound recording and reproduction - brought new opportunities for musical registration, previously possible only by the musical notation (with its limitations, as the expressiveness of performance is not registered by notation), and for musical composition - the need arose for techniques capable of combining composition, arrangement, orchestration, interpretation, improvisations, timbres, and performance into an immutable sonic whole (BURGESS, 2014, p. 1), or an immutable "sound object"[11]. In Burgess' definition:

> Music production is the technological extension of composition and orchestration. It captures the fullness of composition, its orchestration, and the performative intentions of the composer(s). In its precision and inherent ability to capture cultural, individual, ambient, timbral, and interpretive subtleties along with intonation, tempo, intention, and meaning (except when seeking the amorphous), it is superior to written music and oral traditions. Music-making is representational and an art in itself (BURGESS, 2013, p. 5, our translation).

According to the Grammy Award Eligible Credit Definitions (RECORDING ACADEMY, 2019), the music producer is the person responsible for creative, technical, and aesthetic decisions that meet the goals of the artist and the copyright owner of the sound recording in the creation of musical content, often being considered, when this is no longer the case, another member of the musical group, with the same or even greater importance as the musicians. The producer may execute, direct performances, choose final takes or versions and oversee the selection of music, musicians, singers, arrangers, studios, etc. They are also responsible for performing or supervising the mixing, mastering[12] and overall quality control of a musical recording.

[11] A term created by Pierre Schaeffer to refer to an audio excerpt with a unit of sound information whose imagery reference is latent or non-existent (MELO; PALOMBINI, 2006).

[12] " Mastering is the final stage of optimizing the recorded material as it is transferred to the format(s) that will be used in the manufacturing process" (*Mastering is the final stage of optimization of the recorded material while transferring it to the format(s) that will be used in the manufacturing process*) (BURGESS, 2014, p. 48, our translation).

Mixing can be defined as using, in a creative and sometimes intuitive way, techniques and tools to mix, shape and equalize the sound of one or more audio channels with content from different sound sources to achieve a specific aesthetic goal (ARAÚJO, 2015). The use of mixing creatively, as well as its technical aspects, can be evidenced in a clearer and more detailed way:

> Mixing music is related to processing recorded musical performances. The goal of this process can be to make the recording sound natural and realistic as if you were in the room when the musicians performed. It can also be used to dramatically alter the sonic character of the recording, creating a very different soundscape that may not even be possible to achieve in real life. To do this, the mixing engineer has a wide variety of analog and digital tools. These tools are called signal or effects processors. [...] Mixing engineers can therefore use signal processing for other than purely technical reasons. Signal processing can be used in aesthetic and creative ways to make things sound bigger, more passionate, and more emotional. Even though the original signal can be heavily skewed or distorted in the process, making it sound unnatural or with inferior audio quality, it is often considered desirable. [...] When mixing music, mixing engineers sometimes use signal processing to elicit a specific emotional impact on the listener. For example, a vocal track may be mixed with much reverb and delay to induce a dreamlike or melancholic emotion (OLSSON, 2015, p. 2, our translation).

According to Burgess (2013), music production is preceded by the pre-production stage and followed by the post-production stage:

a) Pre-production: preparatory decision-making phase for selecting, organizing, and refining musical material;

b) Production: prepare (choose and position) microphones, instruments, headphones, effects (delay, reverb, etc.), equalizers, and compressors (pre-mixing) and perform the recording sessions, all based on aesthetic choices initiated in the previous stage;

c) Post-production: stage of the mix, which consists of balancing and optimizing the components of the production for maximum musical impact and perceptual clarity of the parts, using resources such as equalization, compression, panning[13], compression, limiting, expansion, gating, reverberation, delays and other effects to optimize the sounds, increase their impact and ensure that they occupy their own space in the audio spectrum. And finally, mastering is the preparation of a single media or digital file with the union of all other recordings that make up a musical piece (music) or a set of them (album).

[13] " Panning left or right positioning of sounds between speakers " (*Panning, the left/right placement of sounds between the speakers*) (GIBSON, 2005, p. 22, our translation).

DOI: https://doi.org/10.47519/eiaerh.v8.2022.ID414

Some of these steps are commonly performed by different professionals specialized in specific functions. But "With the ubiquity of digital audio workstations (DAWs), most producers, since the turn of the century, have been able to record and manipulate audio on the DAW of their choice. This further blurs the distinction between audio engineering and production."[14] (BURGESS, 2013, p. 29, our translation).

It is in the post-production phase that the most appropriate steps for this study are concentrated, especially mixing:

> In the post-production stage, the mixing engineer combines the recordings through mixing and editing to obtain a final mix. Predominantly, the more skilled the mixing engineer, the better the final mix will be in terms of production quality. Audio mixing involves the application of signal processing techniques to each recorded audio track, whereby the engineer manipulates the dynamic (dynamic range balancing and compression), spatial (stereo or surround panning and reverb), and spectral (equalization) characteristics of the source material. Once the final mix has been created, it is sent to a mastering studio, where additional processing is applied so that the musical recording can be distributed for listening in a home or club setting (RONAN; REISS; GUNES, 2018, p. 1, our translation).

Regardless of the specific role performed by the music production professional, keen musical perception is prevalent in performing this type of work.

Music Perception

Mixing is a fundamentally important step in the outcome of recording a piece of music, as David Gibson (2005, p. 17) argues: “Mixing may be only a small part of all that is required to create a great overall recording, yet it is one of the most powerful aspects because the mix can be used to hide weaknesses in other areas.” However, the same author points out that “[...] most people do not differentiate between the individual parts that make up a piece of recorded music. Instead, they listen to an overall 'sound' and rarely separate the mix from the music.” (GIBSON, 2005, p. 1).

To achieve a satisfactory mix, one must first focus on music perception, an innate human ability to “[...] perceive aurally, reflect on, and act creatively upon the music.”(BERNARDES, 2001, p. 75), developed throughout our evolution as a species from sound perception, a defense and protection mechanism to keep us always attentive to events in the surroundings (that is why we can close our eyes but not our ears), and that can also become more refined through training.

[14] *“With the ubiquity of digital audio workstations (DAWs), most producers, since the turn of the century, have been capable of recording and manipulating audio in their workstation of choice. This further blurs the distinction between audio engineering and production.”*

To better understand what musical perception is all about, we turn to a more detailed definition:

> Music perception is a sound perception in the context of music, that is, the ability to perceive sound waves as part of a musical language. Music perception primarily involves sound perception, the ability to identify physical attributes of sound, such as volume, timbre, and pitch. In addition to sound perception, music perception also involves musical elements such as melody (melodic perception), rhythm (rhythmic perception), and harmony (harmonic perception) (MATUNOBU, 2010, p. 22, our translation).

Sound perception occurs, for example, when we are amid a complex soundscape[15] consisting of many sounds of various origins, we are immersed in a large amount of sound information whose source and nature we often do not know. Yet we can perceive specific nuances of these sounds - even if we cannot understand them in depth - such as intensity, pitch, timbre, reverberation, etc. These are the perceptual aspects of sound. In music, one of the fundamental elements of sound perception concerns the mental processing, by hearing, of elementary aspects of sound that describe psychoacoustic[16] characteristics of the material heard" (FORNARI, 2010, p. 12, our translation).

When we are immersed in a disturbed soundscape and perceive sound information all around us coming from the most diverse sources,

> Our ears receive, translate, and send all this sound information to the brain via the auditory nerve through electrical signals. Although this perceptual information is entangled in the two receiving channels, which are the ears, we are able, to some extent, to voluntarily focus our attention on a single conversation, as well as move our attention from one sound source to another according to our interest, and disregard the rest. If someone calls our name in this tumultuous sound environment, especially if we notice that it is a familiar voice, our attention is immediately and involuntarily shifted to this person (FORNARI, 2010, p. 21, our translation).

[15] A concept popularized by R. Murray Schafer (2001, p. 24, our translation): "A soundscape consists of heard events rather than seen objects," and can be divided into fundamental sounds, the notes that identify the scale or tonality of a song or the sounds created by geography and climate (water, wind, birds, insects, animals); signals, sounds highlighted and consciously heard as acoustic warning resources (bells, whistles, horns, sirens); and sound marks, sounds unique to a community that possesses certain qualities that make them especially meaningful or noticeable to the people of that place (melting glaciers, erupting volcanoes, boiling sulfur fields) (SCHAFER, 2001).

[16] "Psychoacoustic features occur at sufficiently small time intervals before the formation of a memorization model of sound information (thus, there is no distinction between sound and musical psychoacoustic aspects). Such aspects are associated with a time interval known in psychoacoustics as the auditory persistence interval, considered to be around 0.1s in duration. Distinct sound events separated by time intervals shorter than the auditory persistence interval are perceived as a single sound event" (FORNARI, 2010, p. 10, our translation).

DOI: https://doi.org/10.47519/eiaerh.v8.2022.ID414

According to Gestalt theory (LERDAHL; JACKENDOFF, 1996; TENNEY; POLANSKY, 1980), there are four basic principles of sound object identification in music (FORNARI, 2010, p. 26-27, our translation):

1. Structuring: understanding a set of different events as a single structure. E.g., instruments, melody, harmony, rhythm, etc.;
2. Segregation: noticing an event that stands out about the others. E.g., the melody of a solo instrument;
3. Pregnancy: to identify the simplest and most regular structures first and clearly. E.g., simple rhythmic patterns (as opposed to polyrhythmic ones);
4. Constancy: to perceive continuity in the variations between consecutive events and understand them as belonging to the same context. E.g., a car passing by with music playing.

With this enhanced capacity for sound perception, R. Murray Schafer (2001, p. 25) states that: “What the soundscape analyst needs to do, in the first place, is to discover its significant aspects, those sounds that are important because of their individuality, quantity, or preponderance.” And this discovery must also be performed by the music producer when mixing the parts that make up a recorded piece of music.

Schafer argues that one way to train in sound perception is what he metaphorically calls “ear cleaning”:

> You start by listening to sounds. The world is full of sounds that can hear everywhere. The most obvious kinds of sounds are also the least heard, which is why the ear-cleaning operation focuses on them. Some students have cleaned their ears so much to hear the sounds around them that they can later analyze them. When the analysis process is accurate, it is possible to reconstruct or imitate a sound that is heard synthetically. This is where ear-cleaning gives way to auditory training (SCHAFER, 1991, p. 103-104, our translation).

Given the above, one can notice the importance of sound perception in music since the musician is also immersed in a sound landscape that, in this case, refers to musical performance, being even more significant in terms of communication between agents in the case of group performance. Thus, a web application was developed in which the user can perform sound perception training and learn some tools used by music producers, as presented in the next section.

A computational model of music perception

The best way to learn to listen is by listening,[17] as Schafer argues in his book The Thinking Ear, 1991. Therefore, a web application whose inputs and feedback are given using sounds presents a high potential for the training of sound perception, in the same way as audio games, which present some advantages in music education due to their emphasis on sound resources (music, sound effects, and voices) and the decrease or even absence of visual resources, as pointed out by Rovithis, Mniestris, and Floros:

> In audio games [AGs], players must focus on auditory stimuli to understand and perform game tasks. Reducing or excluding visual information can enhance the acquisition of skills such as memory and concentration. Moreover, GAs can introduce everyone, even non-musicians, to musical concepts and principles, serving as platforms on which players experiment and realize their ideas. Thus, GA design can play an innovative role in research and education, especially in curricula related to music and sound studies (ROVITHIS; MNIESTRIS; FLOROS, 2014, p. 1, our translation).

Given these possibilities, we present a simple prototype of a web application for training in sound perception and music production, with which the user will be able, through input by voice commands (making the app accessible to the visually impaired), to mix the music being played, by:

- Adding and removing instruments, named as a bass drum, box, tom, cymbals, bass, synth, arpeggio, melody, and effects;
- Switching audio effects on and off, named as chorus, compressor, delay, distortion, flanger, high-pass filter, low-pass filter, reverb, and tremolo;
- Change panning to center (mono), left or right.

The prototype was developed in the engine (software optimized for creating games) Unity[18]. Its choice was due to the possibility of integration with the Fmod[19] middleware[20] which allowed the activation or deactivation of musical instruments and sound effects according to the user's inputs, as well as changing the panning of the music.

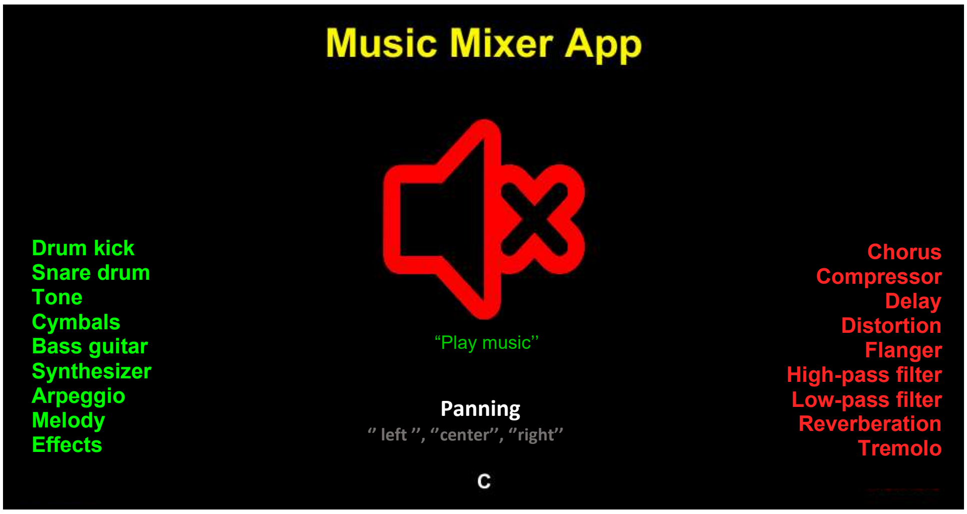
o keep the user of the prototype app focused on the sounds without possible dispersion because of too many visuals, we chose to use input by voice commands, which was made

[17] Or even “listen by listening”, since to listen is to listen attentively, consciously, according to Houaiss (2009, entry “listen”): 1) be aware of what you are hearing; 2) be attentive to listen, pay attention to; 3) make an effort to listen clearly.
[18] Unity's official website: https://unity.com/pt. Access: 10 Jan. 2022.
[19] Official Fmod website: https://www.fmod.com/. Access: 10 Jan. 2022.
[20] Middleware is computer software that provides services to application software beyond those available from the operating system.

DOI: https://doi.org/10.47519/eiaerh.v8.2022.ID414

possible by the WebGL Speech,[21] plugin, written in the C# programming language (the same used in Unity), which allows speech recognition by web browsers and converts the user's speech into text (speech-to-text), more specifically into a string variable (which stores words), that can be compared or manipulated to transform the user's speech commands into actions in the application. Thus, the user can, for example, pronounce the name of a musical instrument or a sound effect, as shown in Figure 1, to turn it off if it is on or on if it is off, which was possible with the creation of a binary variable of the bool type (which stores only two possible values (true or false) to know if the instrument in question is on or off, and thus perform the action commanded by the user. When an instrument or audio effect is on, its name appears green on the screen, and when it is off, the name turns red.

Figure 1 - The prototype's user interface, with the instrument names on the left, audio effects on the right, and the music playback indicator (speaker icon) and panning position in the center.



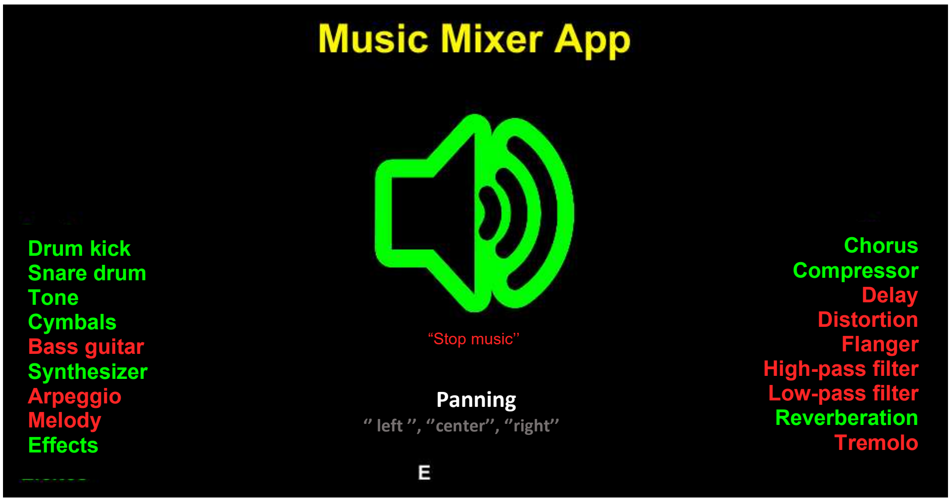
Source: Prepared by the authors

The same premise is used for the user to start or stop playing the music track, which is indicated by a speaker icon in the center of the screen, which turns red when the music is not playing (Figure 1) and turns green when the music is playing (Figure 2). To play the music, the user must say “play music”; to stop it, the user must pronounce “stop the music”.

[21] Available at: https://assetstore.unity.com/packages/tools/audio/webgl-speech-105831. Access: 10 Jan. 2022.

DOI: https://doi.org/10.47519/eiaerh.v8.2022.ID414

Figure 2 - The green speaker icon in the center indicates that music is playing, and the color of the instrument names and audio effects indicate whether they are active (green) or inactive (red).



Music Mixer App

Drum kick
Snare drum
Tone
Cymbals
Bass guitar
Synthesizer
Arpeggio
Melody
Effects

"Stop music"

Panning
" left ", "center", "right"
E

Chorus
Compressor
Delay
Distortion
Flanger
High-pass filter
Low-pass filter
Reverberation
Tremolo

Source: Prepared by the authors

To change the panning, the user can say "center", "left," or "right" so that the music will play, respectively, in both speakers, in mono (Figure 1), only in the left speaker (Figure 2) or only in the right speaker, and a letter (C, E or D) will appear at the bottom of the center of the screen, below the word "Panning".

By default, the application was programmed to recognize words pronounced in Portuguese, more specifically, Brazilian Portuguese (PT-BR). However, for some unknown reason, the code used is not working as expected. For example, when the user accesses the application through a browser configured in a language other than PT-BR, the prototype does not work since the application only recognizes words pronounced in the PT-BR language. Thus, the user needs to manually change the browser's language setting to PT-BR for the model to work correctly.

Furthermore, some problems occurred in the tests performed with speech recognition, which needed to be corrected for the pronunciation of English words when transcribing them into Portuguese. Therefore, for certain string variables to be compared and the intended result to be possible, it was necessary to make adaptations or the spelling of English words, such as "chorus", spelled as "corus", and also adapted to "chorus", which are the forms that the WebGL Speech plugin's speech recognition usually transcribes the pronunciation of "chorus". The speech recognition could hardly identify the pronunciation of the word "flanger" and instead understood the word "creak", so these two options were included to turn this effect on or off when the user pronounces its name.

When the user pronounces any of the keywords, being the commands to turn instruments or effects on or off, the respective parameters created in Fmod are changed, which causes the effects or instruments to be turned on or off. In other words, the parameters were created within Fmod but are manipulated by Unity according to the voice command inputs captured by the speech recognition algorithm of the WebGL Speech plugin, which converts the spoken words into text (speech-to-text). So these words are compared to variables of type string, and if the comparison is effective, a certain previously programmed action occurs. This synergy between Unity, Fmod, and WebGL Speech enabled the development of the web application prototype.

The prototype of this application is available for testing on itch.io[22] (a site for hosting and distribution, paid or free, of independent games) and can be accessed through the link https://leopassos.itch.io/musicmixer.

Final considerations

In some presentation sessions of Music Mixer, as well as in some playtests,[23] people expressed enthusiasm and fun with the application. Some improvements need to be made so that the proposal of offering music production education and sound perception training is more effective and comes even closer to real mixing and music production practice, even if within certain limits, since not all mixing techniques and tools are available in the application. The use of speech recognition inputs brings some drawbacks, such as the delay between the input and the action and feedback by the application; the problems due to the language program in the speech recognition and the one set in the browser; the inaccuracy of the speech recognition system, which often fails to correctly capture the user's pronunciation, especially when the user is not using headphones, and the sound emitted by the speakers sometimes ends up disturbing the speech recognition of the application; and the limitations of speech inputs, being less precise and dynamic than other types of information, such as those performed by mouse or touch screen, which allow a greater variety of commands and actions by the user.

As future possibilities, we intend to allow the user to: include audio samples captured in real time; change the tempo of the music; make touchscreen inputs in a mobile version to

[22] *"itch.io is an open marketplace for independent digital creators with a focus on independent video games. It's a platform that enables anyone to sell the content they've created. As a seller you're in charge of how it's done: you set the price, you run sales, and you design your pages. It's never necessary to get votes, likes, or follows to get your content approved, and you can make changes to how you distribute your work as frequently as you like*." Available at: https://itch.io/.

[23] Some of the playtests were recorded and are available at: https://youtu.be/HrxLHzuhg3w. Access: 10 Jna. 2022.

DOI: https://doi.org/10.47519/eiaerh.v8.2022.ID414

increase the possibilities of use, including the option of adding audio effects to specific instruments, and not to the music as a whole, as the current case of the prototype; and allow volume control and panning of each instrument. With these implementations, the user can accomplish something very close, if not identical, to the mixing of the individual parts that make up a recorded piece of music. However, it is a challenge to carry out these implementations without abandoning the possibility of using voice commands so that the application remains accessible to the visually impaired, as well as the possibilities of using speech synthesis, with the text-to-speech system included in the WebGL Speech plugin.

As future developments of works derived from this one, aiming at the development of a more sophisticated application and with more elaborate possibilities of use, we intend to use the methodology of action research (TRIPP, 2005), which consists of 1) development, 2) tests, 3) improvements, 4) collection of results and 5) restarting the process from step 1. Stages 2 and 4 it is considered to collect data through playtests followed by the users filling out an online report, with due authorization from the Research Ethics Committee (CEP) of the State University of Campinas (Unicamp).

REFERENCES

ARAÚJO, D. V. G. **Uma breve história da mixagem**: Origem, técnicas, percepção e futuros avanços. Campinas, 2015. Dissertação (Mestrado em Música) – Instituto de Artes, Universidade Estadual de Campinas, São Paulo, 2015. Available at: https://revistas.nics.unicamp.br/revistas/ojs/index.php/nr/article/view/190. Access: 06 May 2021.

BENWARD, B.; SAKER, M. **Music in theory and practice**: v. 1. 8. ed. New York: McGraw-Hill, 2009.

BERNARDES, V. A percepção musical sob a ótica da linguagem. **Revista da Abem**, v. 9, n. 6, p. 73-82, set. 2001. Available at: www.abemeducacaomusical.com.br/revistas/revistaabem/index.php/revistaabem/article/view/444. Access: 05 Nov. 2021.

BURGESS, R. J. **The art of music production**: The theory and practice. 4. ed. New York: Oxford University Press, 2013.

BURGESS, R. J. **The history of music production**. New York: Oxford University Press, 2014.

CASE, A. U. **Mix Smart**: Pro audio tips for your multitrack mix. Oxford: Focal Press, 2011.

FORNARI, J. Percepção, cognição e afeto musical. *In*: KELLER, D. (org.). **Criação musical e tecnologias**: Teoria e prática interdisciplinar. Goiânia: Anppom, 2010. Available at: www.anppom.com.br/ebooks/index.php/ pmb/catalog/book/2. Access: 21 June 2021.

GIBSON, D. **The art of mixing**: A visual guide to recording, engineering, and production. 2. ed. Boston: Thomson Course Technology, 2005.

HOUAISS, A [Instituto]. **Houaiss Eletrônico**. Versão 3.0. Rio de Janeiro: Objetiva, 2009.

HURON, D. Affect induction through musical sounds: an ethological perspective. **Phil. Trans. R. Soc. B**, v. 370, n. 1664, mar. 2015. Available at: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2014.0098. Access: 10 Feb. 2021.

HURON, D. Understanding Music-related emotion: Leslons from Ethology. *In*: PROC. INTERN. CONF. ON MUSIC PERCEPTION AND COGNITION, 12.; TRIENNIAL CONF. OF THE EUROPEAN SOC. FOR THE COGNITIVE SCIENCES OF MUSIC, 8., 2012, Thessaloniki. **Anais** […]. Thessaloniki, Greece, 2012.

LERDAHL, F.; JACKENDOFF, R. S. **A generative theory of tonal music**. 3. ed. London: MIT Press, 1996.

MATUNOBU, Y. **Desenvolvimento de software educativo para treinamento em percepção musical**. 2010. Monografia (Trabalho de Conclusão de Curso em Ciência da Computação) – Fundação de Ensino Eurípides Soares da Rocha, Centro Universitário Eurípides de Marília, São Paulo, 2010.

MELO, F.; PALOMBINI, C. O objeto sonoro de Pierre Schaeffer: Duas abordagens. *In*: XVI ANPPOM, 16., 2006, Brasília. **Anais** […]. Brasília, 2006. Available at: https://antigo.anppom.com.br/anais/anaiscongresso_anppom_2006/CDROM/COM/07_Com_TeoComp/sessao04/07COM_TeoComp_0404-173.pdf. Access: 24 May 2021.

OLSSON, E. **Aesthetic signal processing in music production**: Is the intended emotional response achieved? Lulea. 2015. Monografia (Trabalho de Conclusão de Curso em Engenharia de Áudio) – Department of Arts, Communication and Education, Lulea University of Technology, 2015. Available at: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1018575&dswid=-3321. Access: 23 Feb. 2021.

RECORDING ACADEMY. **Producers & Engineers Wing, Technical Guidelines**. Producer Grammy Award Eligibility Crediting Definitions, March 01, 2019. Available at: www.grammy.com/sites/com/files/producer_definitions_final_03_01_2019.pdf. Access: 17 June. 2021.

RONAN, D.; REISS, J. D.; GUNES, H. An empirical approach to the relationship between emotion and music production quality. **ArXiv**, mar. 2018.

ROVITHIS, E.; MNIESTRIS, A.; FLOROS, A. Educational audio *game* design: sonification of the curriculum through a role-playing scenario in the audio *game* 'Kronos'. *In*: AM 2014, 9., 2014, New York. **Anais** […]. New York, NY, USA, 2014.

SCHAFER, R. M. **A afinação do mundo**. São Paulo: Ed. Unesp, 2001.

SCHAFER, R. M. **O ouvido pensante**. São Paulo: Fundação Editora da Unesp, 1991.

TENNEY, J.; POLANSKY, L. Temporal Gestalt perception in music. **Journal of Music Theory**, Autumn, v. 24, n. 2, p. 205-241, 1980. Available at: https://www.jstor.org/stable/843503. Access: 19 Feb. 2021.

TRIPP, D. Pesquisa-ação: Uma introdução metodológica. **Educação e Pesquisa**, São Paulo, v. 31, n. 3, p. 443-466, set./dez. 2005. Available at: http://educa.fcc.org.br/pdf/ep/v31n03/v31n03a09.pdf. Access: 21 Sept. 2021.

TURINO, T. **Music as social life**: The politics of participation. Chicago: The University of Chicago Press, 2008.

ABOUT THE AUTHORS

Leonardo Porto PASSOS

State University of Campinas (UNICAMP), Campinas – SP – Brasil. Master's student at the Graduate Program in Music (PPGM) at the Institute of Arts (IA).

José FORNARI

State University of Campinas (UNICAMP), Campinas – SP – Brasil. Career Researcher Pq at CPG/DM/IA. Doctoral degree in Electrical Engineering (UNICAMP).

Processing and publication by the Editora Ibero-Americana de Educação.

Reviewing, formatting, standardization and translation.